



Comparing different sampling schemes for approximating the integrals involved in the semi-Bayesian optimal design of choice experiments

J. Yu, P. Goos and M. Vandebroek

DEPARTMENT OF DECISION SCIENCES AND INFORMATION MANAGEMENT (KBI)

# Comparing different sampling schemes for approximating the integrals involved in the semi-Bayesian optimal design of choice experiments

**Jie Yu**

Faculty of Business and Economics, Katholieke Universiteit Leuven, Naamsestraat 69, B-3000 Leuven, Belgium.

Email: Jie.Yu@econ.kuleuven.be

Tel: +32 (0)16 32 69 62

Fax: +32 (0)16 32 66 24

**Peter Goos**

Faculty of Applied Economics, Universiteit Antwerpen, Prinsstraat 13, B-2000 Antwerpen, Belgium.

Email: Peter.Goos@ua.ac.be

Tel: +32 (0)3 220 40 59

Fax: +32 (0)3 220 48 17

**Martina Vandebroek**

Faculty of Business and Economics & Leuven Statistics Research Centre, Katholieke Universiteit Leuven, Naamsestraat 69, B-3000 Leuven, Belgium.

Email: Martina.Vandebroek@econ.kuleuven.be

Tel: +32 (0)16 32 69 75

Fax: +32 (0)16 32 66 24

# Comparing different sampling schemes for approximating the integrals involved in the semi-Bayesian optimal design of choice experiments

## Abstract

In conjoint choice experiments, the semi-Bayesian  $D$ -optimality criterion is often used to compute efficient designs. The traditional way to compute this criterion which involves multi-dimensional integrals over the prior distribution is to use Pseudo-Monte Carlo samples. However, other sampling approaches are available. Examples are the Quasi-Monte Carlo approach (randomized Halton sequences, modified Latin hypercube sampling and extensible shifted lattice points with Baker's transformation), the Gaussian-Hermite quadrature approach and a method using spherical-radial transformations. Not much is known in general about which sampling scheme performs best in constructing efficient choice designs. In this study, we compare the performance of these approaches under various scenarios. We try to identify the most efficient sampling scheme for each situation.

Keywords: conjoint choice design, Pseudo-Monte Carlo, Quasi-Monte Carlo, Gaussian-Hermite quadrature, spherical-radial transformation

## 1 Introduction

Conjoint choice experiments have become popular for assessing consumers' preferences about the characteristics of products or services in fields like marketing, econometrics and transportation. A popular econometrics model for analyzing the data from such kinds of experiments is the logit model (McFadden 1974, 1978; Ben-Akiva and Lerman 1985; Brownstone and Train 1999). It has been shown that constructing an efficient choice design for such a model is crucial for improving the statistical inference about the quantities of interests (e.g. Huber and Zwerina 1996; Sándor and Wedel 2001; Vermeulen et al. 2008, Yu et al. 2008a, 2008c).

To construct an efficient choice design, one has to deal with the fact that probabilistic choice models are nonlinear in the parameters. This implies that researchers need to assume values for the parameters before constructing the designs (Atkinson and Donev 1992; Atkinson and Haines 1996; Sándor and Wedel 2001; Dror and Steinberg 2006). In recent years the semi-Bayesian approach, introduced in the marketing literature by Sándor and Wedel (2001), has been used widely for choice experiments (Bliemer et al. 2008; Kessels et al. 2006, 2008; Vermeulen et al. 2008; Yu et al. 2008a, 2008b, 2008c). This approach considers a range of possible values of the parameters by assuming a prior distribution for them in the design construction. Sándor and Wedel (2001) showed the benefits of this approach compared to the locally optimal approach and concluded that taking into account the prior uncertainty in the design stage leads to designs that are highly efficient even if the prior information is not well specified.

The traditional way to compute the semi-Bayesian design criterion which involves a multi-dimensional integral over the prior distribution, is to use Pseudo Monte Carlo (*PMC*) samples. These samples are chosen randomly from the prior distribution. This method is easy to implement but is very likely to lead to a large variability in the result when the number of random draws is small. The reason is that the *PMC* draws are chosen independently of each other and the scattering might be large in some areas of the integration domain while very small in other areas. As a result, different sets of random *PMC* draws are likely to produce different coverages of the distribution space (Bliemer et al. 2008).

To circumvent the lack of stability of the *PMC* approach, researchers often use a large sample of draws to approximate the design criterion which is very time consuming. Sándor and Wedel (2001), Kessels et al. (2006), Vermeulen et al. (2008) and Yu et al. (2008a) all used 1000 *PMC* draws for constructing semi-Bayesian designs for conditional logit models. However, the large sample of *PMC* draws makes the application of semi-Bayesian design criteria for constructing realistic choice experimental designs computationally demanding as there might be millions of integrals that need to be evaluated when searching for the best design.

In recent years, some researchers started to use systematic draws instead of random draws to compute the integrals. Samples from the uniform distribution are obtained using deterministic numbers rather than computer-generated pseudo-random numbers. These deterministic numbers are called Quasi-Monte Carlo (*QMC*) samples. Quasi-random samples are more evenly scattered throughout the domain where the integral is calculated, and therefore improve the accuracy of the approximation of integrals. For this reason, they have received a lot of attention for solving different problems in the literature. Hickernell et al. (2000) have proposed extensible shifted lattice points (*ESLP*) and used them for computing multivariate normal probabilities and evaluating multi-dimensional integrals arising in physics. Sándor and András (2004) have compared several *QMC* methods for estimating multivariate normal probabilities in econometrics and concluded that the *ESLP* with Baker's transformation performs best. Other work, which employs *QMC* for

computing the integrals arising in econometric models, was done by Hess et al. (2006) who have applied the modified Latin hypercube sampling (*MLHS*) to estimate a mixed logit model, Bhat (2001, 2003) and Train (2000, 2003) who employed Halton sequences and a randomized version of them when estimating probabilities for logit models with random coefficients. Sándor and Train (2004) applied  $(t, m, s)$ -nets and Halton draws for an application of maximum simulated likelihood estimation of a mixed logit model.

Also in the context of experimental designs, the use of systematic draws has been studied by many authors. Sándor and Wedel (2002, 2005) utilized samples based on orthogonal arrays and shifted good lattice points for constructing locally mixed logit designs, respectively. Recently Kessels et al. (2008) constructed semi-Bayesian conditional logit designs using a small set of systematic draws based on the theory of minimum potential energy. Their approach was extended by Yu et al. (2008b) for semi-Bayesian mixed logit designs. Gotwalt et al. (2008) modified the approach of Monahan and Genz (1997) to compute a semi-Bayesian design criterion. The integral related to the multivariate normal prior distribution was reparameterized and decomposed into a radial integral and an integral over the surface of the unit hypersphere. They showed that the proposed sampling approach is quite efficient in approximating a semi-Bayesian design criterion. Bliemer et al. (2008) compared *PMC* samples, some *QMC* samples and samples obtained using a Gaussian-Hermite quadrature approach. They showed the superiority of the Gaussian-Hermite quadrature method in the context of experimental choice designs. The relative performance of the Gaussian-Hermite quadrature approach compared to other recently developed approaches like *ESLP* with Baker's transformation and the approach proposed by Gotwalt et al. (2008) in evaluating the semi-Bayesian design criterion for choice models has not been studied yet and will be investigated in this paper.

With small samples of intelligent draws, many problems which remained unsolved for a long time can be tackled now. A typical example is the construction of optimal experimental designs for nonlinear models using the true Bayesian design criteria where the expected posterior covariance matrix is considered. All authors mentioned before studied semi-Bayesian design criteria based on the Fisher information matrix (*FIM*). The underlying idea is that the inverse of the *FIM* is an asymptotic approximation to the posterior covariance matrix of the parameter estimates when the sample size is large. With the traditional *PMC* approach where a large number of draws is required, finding the semi-Bayesian design based on the *FIM* is feasible but computationally demanding. It is however impossible to construct optimal designs based on the covariance matrix of the posterior distribution using the *PMC* approach. Using the results of this paper, however, Yu et al. (2008c) were able to construct Bayesian designs for conditional logit models based on the expected posterior covariance matrix.

The literature shows that the *ESLP* with Baker's transformation, the Gaussian-Hermite quadrature and the approach proposed by Gotwalt et al. (2008) are the most promising sampling approaches. Each of them has emerged as the best sampling approach in some studies. However,

these three approaches have not been compared with each other. In this paper, we study the relative performance of these three sampling approaches in generating efficient experimental choice designs. In addition, we also include three other well-known sampling approaches (*MLHS*, randomized Halton sequences and *PMC*) in the comparison.

## 2 Sampling Methodologies

In this section, we present six sampling approaches for approximating multi-dimensional integrals. These include *PMC* samples, *QMC* samples (randomized Halton sequences, *MLHS* and *ESLP* with Baker's transformation), the Gaussian-Hermite quadrature, and the method based on a spherical-radial transformation.

Suppose that  $\boldsymbol{\beta}$  follows a multivariate normal distribution,  $\boldsymbol{\beta} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ . The expectation of a function  $g(\boldsymbol{\beta})$  with respect to the distribution of  $\boldsymbol{\beta}$ ,  $\pi(\boldsymbol{\beta})$ , is expressed as

$$\begin{aligned} E[g(\boldsymbol{\beta})] &= \int_{\mathbb{R}^p} g(\boldsymbol{\beta}) \pi(\boldsymbol{\beta}) d\boldsymbol{\beta}, \\ &= \int_{\mathbb{R}^p} g(\boldsymbol{\beta}) (2\pi)^{-\frac{p}{2}} |\boldsymbol{\Sigma}|^{-\frac{1}{2}} e^{-\frac{1}{2}(\boldsymbol{\beta}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{\beta}-\boldsymbol{\mu})} d\boldsymbol{\beta}. \end{aligned} \tag{1}$$

To take draws from a multivariate normal distribution, one often transforms the multivariate normal distribution into univariate standard normal distributions. The transformation is based on a Cholesky decomposition of the covariance matrix  $\boldsymbol{\Sigma} = \mathbf{D}\mathbf{D}'$ . This leads to  $\boldsymbol{\beta} = \boldsymbol{\mu} + \mathbf{D}\mathbf{z}$ , where the vector  $\mathbf{z}$  has elements drawn from independent standard normal distributions. As

$$(\boldsymbol{\beta} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\boldsymbol{\beta} - \boldsymbol{\mu}) = \mathbf{z}' \mathbf{D}' (\mathbf{D}')^{-1} \mathbf{D}^{-1} \mathbf{D} \mathbf{z} = \mathbf{z}' \mathbf{z}, \tag{2}$$

and

$$d\boldsymbol{\beta} = |\mathbf{D}| d\mathbf{z} = |\boldsymbol{\Sigma}|^{\frac{1}{2}} d\mathbf{z}, \tag{3}$$

equation (1) becomes

$$\begin{aligned}
E[g(\boldsymbol{\beta})] &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(\boldsymbol{\mu} + \mathbf{D}\mathbf{z}) (2\pi)^{-\frac{p}{2}} \prod_{k=1}^p e^{-\frac{z_k^2}{2}} dz_1 \cdots dz_p, \\
&= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(\boldsymbol{\mu} + \mathbf{D}\mathbf{z}) \prod_{k=1}^p \phi(z_k) dz_1 \cdots dz_p,
\end{aligned} \tag{4}$$

where  $\phi(z_k) = (2\pi)^{-\frac{1}{2}} e^{-\frac{z_k^2}{2}}$  is the density function of a univariate standard normal distribution.

To approximate the integral in (4),  $N$  draws are taken from the corresponding distribution  $\phi(z_k)$  for each of the  $p$  parameters. Let  $\Phi(z_k)$  be the standard cumulative distribution function. The  $r$ th draw for the  $k$ th parameter,  $z_k^r$ , is then computed as  $z_k^r = \Phi^{-1}(x_k^r)$ , where  $x_k^r, r = 1, \dots, N$  is a set of points randomly taken from a uniform distribution on the interval  $[0, 1)$ . For  $p$  dimensions, the draws are taken from the uniform distribution on  $[0, 1)^p$  and (4) is approximated by

$$E[g(\boldsymbol{\beta})] = \frac{1}{N} \sum_{r=1}^N g(\boldsymbol{\mu} + \mathbf{D}\Phi^{-1}(\mathbf{x}^r)) \tag{5}$$

where  $\Phi^{-1}(\mathbf{x}^r) = [\Phi^{-1}(x_1^r), \dots, \Phi^{-1}(x_p^r)]$ , the draws  $(\mathbf{x}^r)_{r=1}^N$  can be obtained either by a *PMC* method or by a *QMC* method. In the *PMC* method,  $(x_k^r), k = 1, \dots, p$ , are obtained using computer-generated pseudo-random numbers. In *QMC* method,  $(x_k^r), k = 1, \dots, p$ , are obtained using number-theoretic methods (e.g., Halton 1960; Korobov 1959). These are explained in detail in the next section.

## 2.1 Quasi-Monte Carlo Sequences

This section begins with one of the best known *QMC* sequences, namely the randomized Halton sequence and then introduces the *MLHS* approach. Finally, the *ESLP* with Baker's transformation is presented.

### 2.1.1 Randomized Halton Sequences

Halton sequences, proposed by Halton (1960), are designed to span the domain of the  $p$ -dimensional unit cube uniformly and efficiently. They are constructed using a deterministic method based on prime numbers. A different prime base is used for each dimension. For example, for the first dimension, the prime number 2 is used, while for the second dimension, the prime number 3 is utilized.

In one dimension, the  $i$ th element in the Halton sequence based on prime  $b$  is generated by first expanding  $i$  in terms of the base  $b$ :

$$i = \sum_{k=0}^m i_k b^k = i_0 b^0 + i_1 b + i_2 b^2 + \dots, \quad (6)$$

where the digits  $i_k$  are in  $\{0, \dots, b-1\}$ , and  $m$  is determined by  $b^m \leq i < b^{m+1}$ . For instance, if base  $b = 2$ ,  $i = 4$ ,  $2^m \leq 4 < 2^{m+1}$  gives us  $m = 2$ . Integer  $i = 4$  is then expressed as  $4 = 0 \times 2^0 + 0 \times 2^1 + 1 \times 2^2$ .

The  $i$ th element of the one dimensional Halton sequence based on prime  $b$  is then obtained as

$$\phi_b(i) = \sum_{k=0}^m i_k b^{-(k+1)} = i_0 b^{-1} + i_1 b^{-2} + \dots \quad (7)$$

Equation (7) produces the uniformly distributed Halton sequence which is uniformly distributed in  $[0, 1)$ . For instance, if  $N = 7$  points need to be computed using base 2, the first element is determined by  $\phi_2(1) = 1 \times 2^{-1} = \frac{1}{2}$ , the second element is determined by  $\phi_2(2) = 0 \times 2^{-1} + 1 \times 2^{-2} = \frac{1}{4}$ , and the rest of the points can be computed similarly:  $\phi_2(3) = \frac{3}{4}$ ,  $\phi_2(4) = \frac{1}{8}$ ,  $\phi_2(5) = \frac{5}{8}$ ,  $\phi_2(6) = \frac{3}{8}$ ,  $\phi_2(7) = \frac{7}{8}$ .

The  $p$ -dimensional Halton sequence is obtained by pairing  $p$  one-dimensional sequences based on  $p$  consecutive prime integers,  $b_1, b_2, \dots, b_p$ . Suppose  $\mathbf{x}_i$  is a  $p$ -dimensional Halton draw constructed by using the first  $p$  prime numbers  $b_1, b_2, \dots, b_p$ , as bases. The  $i$ th point,  $\mathbf{x}_i$ , in the sequence can then be written as

$$\mathbf{x}_i = (\phi_{b_1}(i), \phi_{b_2}(i), \dots, \phi_{b_p}(i)), i = 1, 2, \dots \quad (8)$$

Halton sequences are systematic rather than random, but they can be transformed in a way that makes them random. The transformation process utilized by Bhat (2003) and Train (2003) was based on the two-step procedure introduced by Tuffin (1996). First, for the  $k$ th one-dimensional sequence ( $k = 1, \dots, p$ ), a random draw  $u_k$  is taken from a standard uniform distribution in space  $[0, 1)$ , and added to each element of the sequence. If the resulting element exceeds one, then 1 unity is subtracted from it. The randomized  $p$ -dimensional Halton sequence can then be written as

$$\mathbf{x}_i = (\{\phi_{b_1}(i) + u_1\}, \dots, \{\phi_{b_p}(i) + u_p\}), i = 1, 2, \dots, \quad (9)$$



where the symbol  $\{x\}$  denotes the fractional part of the number  $x$ . By adding the same random number  $u_k$  to each element in dimension  $k$ , the relation between the points in dimension  $k$  and the degree of coverage is not changed.

### 2.1.2 Modified Latin Hypercube Sampling

The *MLHS* proposed by Hess et al. (2006) constructs multi-dimensional sequences by combining randomly shuffled one-dimensional sequences. The  $i$ th point,  $\mathbf{x}_i$ , in the  $p$ -dimensional MLHS sequence can be written as:

$$\mathbf{x}_i = \left( \frac{i-1}{N} + u_1, \dots, \frac{i-1}{N} + u_p \right), \quad i = 1, \dots, N \quad (10)$$

where  $u_k, k = 1, \dots, p$ , is a random number drawn from  $(0, \frac{1}{N})$  with  $N$  the number of points in the one-dimensional sequence. For example, suppose that  $N = 5$ , and that the random number  $u_k$  for the  $k$ th dimension is 0.07, the resulting sequence for the  $k$ th dimension is 0.07, 0.27, 0.47, 0.67, 0.87. It is clear that the distances between adjacent draws are all equal to  $1/N = 0.2$ , satisfying the condition of equal spacing. Finally, the sequence in each dimension obtained in (10) is reordered using a random shuffling permutation. The use of a different shuffling for different dimensions disrupts the cyclical nature of the sequences in different ways.

### 2.1.3 Extensible Shifted Lattice Points

The rank-1 lattice points introduced by Korobov (1959) are another type of low discrepancy set used in *QMC* methods. However, it has the limitation of being of fixed size. More specifically, the construction of the rank-1 lattice points depends on the total number of points,  $N_0$ , in the set. This implies that adding any extra draw destroys the lattice structure. To construct  $N - N_0$  extra lattice points, one has to discard all original  $N_0$  points. Therefore, Hickernell et al. (2000) proposed to construct the infinite shifted lattice points, also called extensible shifted lattice points, which have the attractive property that the number of draws can be increased while retaining the existing points.

According to Hickernell et al. (2000), an infinite rank-1 lattice sequence in base  $b \geq 2$  with generating vector  $\mathbf{q} = [q_1, \dots, q_p]$  and shift  $\mathbf{u} = [u_1, \dots, u_p]$  is defined as

$$\mathbf{x}_i = (\{\phi_b(i)q_1 + u_1\}, \dots, \{\phi_b(i)q_p + u_p\}), \quad i = 0, 1, 2, \dots, b^m - 1 \quad (11)$$

where  $\phi_b(i)$  is given in equation (7) and  $b^m$  is the total number of points. The vector  $\mathbf{q}$  is computed by minimizing some loss function based on a measure of discrepancy or nonuniformity of the lattice (Hickernell et al. 2000). We consider the Korobov type of lattice points where  $\mathbf{q} = (1, q, q^2, \dots, q^{p-1})$  with  $q = 1571$  which is recommended by Hickernell et al. (2000) for samples of size  $2^m$  with  $m = 6, \dots, 12$  and dimensions up to 33. The  $p$ -dimensional shift, with  $u_i$  a random number in  $[0, 1)$ , is meant to randomize the lattice points.

To increase the accuracy of the lattice rule methods, Hickernell et al. (2000) and Sándor and András (2004) utilized the so-called Baker's transformation. All elements in the sequence are transformed by  $x'_{ik} = 1 - |2x_{ik} - 1|$ ,  $i = 0, \dots, b^m - 1$ ,  $k = 1, \dots, p$ , where  $x_{ik} = \{\phi_b(i)q_k + u_k\}$ . The vector  $\mathbf{x}'_i = (x'_{i1}, \dots, x'_{ip})$  is called extensible shifted lattice point with Baker's transformation.

## 2.2 Gaussian-Hermite Quadrature

As shown previously, the *PMC* and *QMC* methods all take an unweighted average of the function values when computing the expectation of  $g(\boldsymbol{\beta})$ , but differ in the way they take draws from the uniform distribution. Similar to the *QMC* methods where systematic draws are used, the Gaussian-Hermite quadrature method constructs systematic draws as well. However, rather than assigning the same weight to all draws as with the *PMC* and *QMC* methods, the Gaussian-Hermite quadrature method gives a specific weight to each draw and computes a weighted sum of the function values.

Prior to explaining how this method works for computing the expectation of  $g(\boldsymbol{\beta})$  over a multivariate normal distribution, we explain how to calculate the integral over the standard normal distribution using the Gaussian-Hermite quadrature approach. In the standard Gaussian-Hermite quadrature, the integrand and its approximation are as follows

$$I[f(x)] = \int_{-\infty}^{\infty} f(x)e^{-x^2}dx \approx \sum_{i=1}^N w_i^N f(a_i^N), \quad (12)$$

where the factor  $e^{-x^2}$  is regarded as a weighting function,  $w_i^N$  and  $a_i^N$  are called standard weights and abscissas and can be computed using Hermite polynomials. If the weighting function is given by a standard normal density,  $\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}$ , the corresponding integration is

$$I[f(t)] = \int_{-\infty}^{\infty} f(t) \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt. \quad (13)$$

Let  $t = \sqrt{2}x$ . Equation (13) then becomes

$$I[f(t)] = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} f(\sqrt{2}x) e^{-x^2} dx \approx \sum_{i=1}^N \tilde{w}_i^N f(\tilde{a}_i^N), \quad (14)$$

where  $\tilde{a}_i^N$  and  $\tilde{w}_i^N$  are the corresponding modified abscissas and weights. Comparing (14) to (12), we notice that the modified abscissas and weights can be obtained from the standard abscissas and weights. That is,  $\tilde{a}_i^N = \sqrt{2}a_i^N$  and  $\tilde{w}_i^N = w_i^N/\sqrt{\pi}$ . Table 1 contains the modified abscissas and weights for  $N$  up to 6.

\*Table 1\*

For the problem that we consider in this paper (computing the expectation of  $g(\boldsymbol{\beta})$  over a multivariate normal distribution), we first follow the procedures given in (1), (2), (3) and (4) to transform the multivariate normal distribution into  $p$  univariate standard normal distributions. The  $N_k$  draws for the  $k$ th standard normal distribution ( $k = 1, \dots, p$ ) are given by the modified abscissas,  $\tilde{a}_{i_k}^{N_k}$ , and the corresponding weights are the modified weight,  $\tilde{w}_{i_k}^{N_k}$ ,  $i = 1, \dots, N_k$ . The weights sum to one in each dimension  $k$ . The number of draws in each dimension,  $N_k$  can be unequal. The  $p$ -dimensional integral of  $g(\boldsymbol{\beta})$  in (4) is then approximated by

$$E[g(\boldsymbol{\beta})] \approx \sum_{i_1=1}^{N_1} \dots \sum_{i_p=1}^{N_p} \left[ \tilde{w}_{i_1}^{N_1} \dots \tilde{w}_{i_p}^{N_p} g(\boldsymbol{\mu} + \mathbf{D}\tilde{\mathbf{a}}_{(i_1, i_2, \dots, i_p)}) \right], \quad (15)$$

where  $\tilde{\mathbf{a}}_{(i_1, i_2, \dots, i_p)} = (\tilde{a}_{i_1}^{N_1}, \tilde{a}_{i_2}^{N_2}, \dots, \tilde{a}_{i_p}^{N_p})$  is a  $p$ -dimensional vector which contains a combination of the abscissas from each of the  $p$  dimensions. The total number of draws is equal to the total number of all combinations of abscissas in all dimensions, that is  $N = \prod_{k=1}^p N_k$ . Obviously, the value of  $N$  grows dramatically as the dimension increases. Therefore, the Gaussian-Hermite quadrature is not suited for practical use in high dimensions.

## 2.3 Method Based on Spherical-Radial Transformation

The last sampling approach that we consider is based on the spherical-radial transformation ( $SR$ ) proposed by Gotwalt et al. (2008). The Gaussian-Hermite quadrature method and the  $SR$  approach both take a weighted sum of the integrand values, but differ in the way they select the draws and the corresponding weights. Rather than working in the original coordinate system, the  $SR$  approach transforms the original system to a radial-spherical coordinate system. The original integration is then transformed into a radial integration and an integration over the surface of the unit sphere. Using the formulas (1), (2) and (3), the expectation of  $g(\boldsymbol{\beta})$  can be expressed as

$$E[g(\boldsymbol{\beta})] = (2\pi)^{-\frac{p}{2}} \int_{\mathbb{R}^p} g(\boldsymbol{\mu} + \mathbf{D}\mathbf{z}) e^{-\frac{1}{2}\mathbf{z}^T\mathbf{z}} d\mathbf{z}. \quad (16)$$

The standard normal vector  $\mathbf{z}$  can then be changed to a radial-spherical coordinate system. Let  $\mathbf{z} = r\mathbf{s}$ , with  $\mathbf{s}^T\mathbf{s} = 1$ , so that  $\mathbf{z}^T\mathbf{z} = r^2$ . The integral in (16) can be decomposed over the  $(p-1)$ -dimensional unit hypersphere  $S^{p-1}$  in  $\mathbb{R}^p$ . Expression (16) then becomes

$$E[g(\boldsymbol{\beta})] = (2\pi)^{-\frac{p}{2}} \int_0^\infty \int_{S^{p-1}} g(\boldsymbol{\mu} + \mathbf{D}r\mathbf{s}) r^{(p-1)} e^{-\frac{r^2}{2}} d\mathbf{s} dr. \quad (17)$$

Let  $r = \sqrt{2x}$ , equation (17) is then written as

$$E[g(\boldsymbol{\beta})] = \frac{\pi^{-\frac{p}{2}}}{2} \int_0^\infty \int_{S^{p-1}} g(\boldsymbol{\mu} + \mathbf{D}\sqrt{2x}\mathbf{s}) x^{\frac{(p-1)}{2}} e^{-x} d\mathbf{s} dx. \quad (18)$$

The outer integral (the radial integral) is approximated with a generalized Gauss-Laguerre quadrature with an abscissa at zero added. Cassity and Hopper (1964) summarized the abscissas,  $a_i$  and weights,  $w_{R_i}$ ,  $i = 0, \dots, N_R$  for use in the following generalized Gauss-Laguerre quadrature formula

$$\int_0^\infty f(x) x^s e^{-x} dx = w_{R_0} f(0) + \sum_{i=1}^{N_R} w_{R_i} f(a_i). \quad (19)$$

Notice that (19) has similar pattern as (18), where  $f(x) = \int_{S^{p-1}} g(\boldsymbol{\mu} + \mathbf{D}\sqrt{2x}\mathbf{s}) d\mathbf{s}$ ,  $s = \frac{p}{2} - 1$ . Based on (19), the outer integral in expression (18) can be approximated as follows

$$E[g(\boldsymbol{\beta})] \approx \frac{\pi^{-\frac{p}{2}}}{2} \left[ w_{R_0} \int_{S^{p-1}} g(\boldsymbol{\mu}) d\mathbf{s} + \sum_{i=1}^{N_R} w_{R_i} \int_{S^{p-1}} g(\boldsymbol{\mu} + \mathbf{D}r_i\mathbf{s}) d\mathbf{s} \right], \quad (20)$$

where  $r_i$ ,  $i = 1, \dots, N_R$  denotes the  $i$ th radius and is equal to  $\sqrt{2a_i}$ . In Table 2, we present the values of the radii  $r_i$  for  $N_R$  up to 4 and the parameter  $s = 1$ . Note that the radial integral is approximated by  $N_R$  radii and their corresponding weights  $w_{R_i}$ ,  $i = 1, \dots, N_R$  and a zero ordinate  $r_0$  and its corresponding weight  $w_{R_0}$ .

\*Table 2\*

The inner spherical integral in (18) is over the surface of the unit sphere  $S^{p-1}$  and can be approximated by the extended simplex integration method proposed by Mysovskikh (1980). It begins with the construction of the  $p+1$  regular points at the vertices of the regular simplex in  $p$  dimensions

$$\mathbf{s}_j^v = (s_{1j}, s_{2j}, \dots, s_{pj}), j = 1, 2, \dots, p+1, \quad (21)$$

where

$$s_{ij} = \begin{cases} -\sqrt{\frac{p+1}{p(p-i+1)(p-i+2)}} & , \quad i < j \\ \sqrt{\frac{(p+1)(p-j+1)}{p(p-j+2)}} & , \quad i = j \\ 0 & , \quad i > j. \end{cases}$$

The extended simplex adds the negatives of these regular vertex points to get  $2(p+1)$  vertex points. In addition, the  $p(p+1)/2$  midpoints of the edges joining all pairs of the original vertex points are projected to the unit hypersphere surface. These projected points are denoted by  $\mathbf{s}_h^m, h = 1, \dots, p(p+1)/2$ . The set  $\{\mathbf{s}^m\}$  which contains the projected points can be constructed as  $\{(\mathbf{v}_i + \mathbf{v}_j)/\|\mathbf{v}_i + \mathbf{v}_j\| = (\mathbf{v}_i + \mathbf{v}_j)/\sqrt{2(p-1)/p}, i < j\}$ . The weight for the regular vertex points and their negatives is given by  $U_p w_{S_v} = U_p [p(7-p)] / [2(p+1)^2(p+2)]$ , where  $U_p = \frac{2(\pi)^{\frac{p}{2}}}{\Gamma(\frac{p}{2})}$  gives the surface area of the unit  $p$ -sphere, while the weight for the projected midpoints and their negatives takes the value  $U_p w_{S_m} = U_p [2(p-1)^2] / [p(p+1)^2(p+2)]$ . These extended simplex points leads to the sphere integral approximation which has order 5. In general, this algorithm generates  $(p+1)(p+2)$  points, but there are some cases in which some points are duplicated. For instance, there are 6 instead of 12 different points for  $p = 2$  and 14 instead of 20 different points for  $p = 3$ . The duplications are clearly shown in Table 3 and Table 4, respectively. For those cases with  $p > 3$ , there are no duplications available.

\*Table 3\*

\*Table 4\*

After the spherical-radial transformation, (20) can be written as:

$$E[g(\boldsymbol{\beta})] \approx w_{R_0} \frac{g(\boldsymbol{\mu})}{\Gamma(\frac{p}{2})} + \sum_{i=1}^{N_R} \frac{w_{R_i}}{\Gamma(\frac{p}{2})} \left\{ w_{S_v} \left[ \sum_{j=1}^{p+1} g(\boldsymbol{\mu} \pm \mathbf{D}r_i \mathbf{s}_j^v) \right] + w_{S_m} \left[ \sum_{h=1}^{\frac{p(p+1)}{2}} g(\boldsymbol{\mu} \pm \mathbf{D}r_i \mathbf{s}_h^m) \right] \right\}. \quad (22)$$

Monahan and Genz (1997) showed that a mixed method which uses fixed quadrature on the radius and randomized spherical integration is superior to the completely randomized method which uses randomized quadrature on the radius and randomized spherical integration. Following their results, we consider a mixed method rather than a completely randomized one. For each nonzero radius component  $r_i$ , we create  $N_M$  random orthogonal matrices,  $\{\mathbf{M}_{ik}\}_{k=1}^{N_M}$  and average the values of the integrand over these orthogonal rotations. Note that  $\mathbf{M}_{ik}$  can be constructed based on the approach proposed by Stewart (1980). With this approach, a  $p \times p$  matrix  $\mathbf{H}$  is first generated with entries  $h_{ij} \sim N(0,1)$ . Then the Householder transformations are used to reduce matrix  $\mathbf{H}$  to a upper triangular matrix  $\mathbf{R}$ . The factorization can be expressed as  $\mathbf{H} = \mathbf{M}\mathbf{R}$  from

which a random orthogonal matrix  $\mathbf{M}$  is obtained. Equation (22) then becomes

$$E[g(\boldsymbol{\beta})] \approx w_{R_0} \frac{g(\boldsymbol{\mu})}{\Gamma(\frac{p}{2})} + \sum_{i=1}^{N_R} \sum_{k=1}^{N_M} \frac{w_{R_i}}{N_M \Gamma(\frac{p}{2})} \left\{ w_{S_v} \left[ \sum_{j=1}^{p+1} g(\boldsymbol{\mu} \pm \mathbf{D} r_i \mathbf{M}_{ik} \mathbf{s}_j^v) \right] + w_{S_m} \left[ \sum_{h=1}^{\frac{p(p+1)}{2}} g(\boldsymbol{\mu} \pm \mathbf{D} r_i \mathbf{M}_{ik} \mathbf{s}_h^m) \right] \right\}. \quad (23)$$

### 3 Simulation Study

In this section, we compare the performance of the six sampling approaches in terms of two measures. The first one is the percentage deviation of the approximated semi-Bayesian  $D$ -criterion value from the true value. The second one is the number of times that the approximated values fall within a certain percentage from the true value.

The semi-Bayesian  $D$ -criterion that we consider in this simulation study is based on the widely used Fisher Information Matrix ( $FIM$ ) for the conditional logit model (McFadden 1974). The conditional logit probability that alternative  $k$  is chosen from a choice set  $s$  is given by

$$p_{ks}(\boldsymbol{\beta}) = \frac{\exp(\mathbf{x}_{ks}' \boldsymbol{\beta})}{\sum_{i=1}^K \exp(\mathbf{x}_{is}' \boldsymbol{\beta})}, k = 1, \dots, K, \quad (24)$$

with  $K$  the number of profiles in each choice set,  $\mathbf{x}_{ks}$  a  $p$ -dimensional vector characterizing the attributes of profile  $k$  in choice set  $s$ , and  $\boldsymbol{\beta}$  a  $p$ -dimensional coefficient vector containing the effects of the different attribute levels on the utility. The  $FIM$ , which is the inverse of the asymptotic covariance matrix of the maximum likelihood estimator, is given by

$$\mathcal{I}(\boldsymbol{\beta} \mid \mathbf{X}) = N \sum_{s=1}^S \mathbf{X}_s' (\mathbf{P}_s - \mathbf{p}_s \mathbf{p}_s') \mathbf{X}_s, \quad (25)$$

where  $N$  is the number of respondents,  $\mathbf{X}_s$  is the design matrix for choice set  $s$ ,  $\mathbf{P}_s = \text{diag}[p_{1s}, p_{2s}, \dots, p_{Ks}]$  and  $\mathbf{p}_s = [p_{1s}, p_{2s}, \dots, p_{Ks}]'$ .

Let  $\pi(\boldsymbol{\beta})$  denote the prior distribution used for constructing the design, then the semi-Bayesian design criterion has the following expression:

$$\phi(\mathbf{X}, \boldsymbol{\beta}) = \int_{\mathbb{R}^p} \det[\mathcal{I}(\boldsymbol{\beta} \mid \mathbf{X})]^{-1/p} \pi(\boldsymbol{\beta}) d\boldsymbol{\beta}. \quad (26)$$

The goal of this study is to compare the performance of the six methods in approximating the criterion value in (26) and to find out which sampling approach is most appealing under different

conditions. To do this, we construct four scenarios (see Table 5) which differ from each other in the number of parameters and in the degree of prior uncertainty. The prior  $\boldsymbol{\beta}$  is assumed to be a multivariate normal distribution,  $\boldsymbol{\beta} \sim N(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$ . For the 4-parameter cases which correspond to scenarios 1 and 2, we define  $\boldsymbol{\mu}_0 = [-1, 0, -1, 0]'$  and  $\boldsymbol{\Sigma}_0 = \sigma^2 \mathbf{I}_4$ , where  $\mathbf{I}_4$  is the 4-dimensional identity matrix and  $\sigma$  takes the values 0.2 and 1. For the 8-parameter cases which correspond to scenarios 3 and 4, we define  $\boldsymbol{\mu}_0 = [-1, 0, -1, 0, -1, 0, -1, 0]'$  and  $\boldsymbol{\Sigma}_0 = \sigma^2 \mathbf{I}_8$ , where  $\mathbf{I}_8$  is the 8-dimensional identity matrix and  $\sigma$  also takes the values 0.2 and 1.

\*Table 5\*

The designs used for assessing the criterion value with different approximation methods were obtained in the following way. For the 4-parameter cases, we used the semi-Bayesian  $D$ -optimal design constructed by drawing a large number of Monte Carlo samples from the multivariate normal prior  $N(\boldsymbol{\mu}_0, \mathbf{I}_4)$ . For the 8-parameter cases, we took the best design from Sándor and Wedel (2001). For each scenario described in Table 5, the criterion value in (26) was approximated using the six different approximations. For scenarios 1 and 2, we set the number of draws  $N=64, 128, 256, 512$  and  $1024$ . For scenarios 3 and 4, we set  $N=128, 256, 512$  and  $1024$ . These numbers were chosen because we use 2 as the base and  $ESLP$  can only take a number of draws  $N$  which is a power of the base. The values of  $N$  used for the randomized Halton and the  $MLHS$  approaches were identical to those for  $ESLP$ . In order to make a fair comparison, we used similar values of  $N$  for the Gaussian-Hermite quadrature and the  $SR$  approaches (see the footnotes in Table 6 and Table 7). For a given type of approximation and a given number of draws  $N$ , 100 approximated criterion values were computed by changing the draws randomly 100 times. These values are then compared to the true one, obtained using 1,000,000 random Monte Carlo draws, in terms of the percentage deviation of the approximated values from the true value. The percentage deviation is defined as

$$PE = \frac{|\phi^A(\mathbf{X}, \boldsymbol{\beta}) - \phi^T(\mathbf{X}, \boldsymbol{\beta})|}{\phi^T(\mathbf{X}, \boldsymbol{\beta})} \times 100, \quad (27)$$

where  $\phi^A(\mathbf{X}, \boldsymbol{\beta})$  and  $\phi^T(\mathbf{X}, \boldsymbol{\beta})$  are the approximated and the true criterion values, respectively.

\*Table 6\*

\*Table 7\*

The results are summarized in Table 6 and Table 7 for scenarios 1 and 2 and scenarios 3 and 4, respectively. The rows labeled  $\overline{PE}$  show us, on average, to what extent the approximated values deviate from the true one for a given number of draws using a given approximation method. The remaining rows in both tables show the  $D(s)$  values, the number of times the approximated values

fall within a certain percentage  $s$  from the true value. For example, in Table 6 with  $N = 64$ , the row with  $D(0.1\%)$  contains a value of 3 for the *PMC* approach, which indicates that using the *PMC* approach, only 3 out of 100 approximated values are within 0.1% of the true value. Note that the draws used in the Gaussian-Hermite quadrature method proposed by Bliemer et al. (2008) are completely deterministic. As a result, we have only one approximated value and no  $D(s)$ -values are available for that method.

Several conclusions can be drawn by comparing the approaches. Not surprisingly, all methods lead to more accurate approximations under small parameter uncertainty than under large parameter uncertainty. This implies that if the researcher is uncertain about the true parameter values, more draws are needed. Secondly, all the sampling approaches based on the systematic draws perform substantially better than the *PMC* approach in most cases we have studied. This implies that in general, with the *PMC* approach, we have to use much larger sample sizes in order to achieve the same precision as with the other approaches. In particular, with small sample sizes which are appealing for practical use, the *PMC* method is not appropriate.

The *SR* approach based on the spherical-radial transformation in general outperforms all other approaches. It is seen that for scenario 1 where the prior uncertainty is relatively small, with only  $N \approx 64$  draws, *SR* approach produces an average percentage deviation of  $\overline{PE} = 0.01\%$ . To achieve the same precision with the *ESLP* approach (the best competitor among the *QMC* sequences),  $N = 1024$  draws are needed. Note that even with  $N = 1024$ , the *PMC*, *MLHS* and the randomized Halton sequences lead to average percentage deviations which are 29 times, 6 times and 4 times higher than the *SR* approach with  $N \approx 64$ . The evaluation in terms of the  $D(s)$  leads to the same conclusion. All the approximated values with the *SR* approach are within 0.01% from the true value. This is also visualized in Figure 1 and Figure 2. Figure 1 presents the boxplot of the percentage deviations when  $N = 64$ . Figure 2 presents the results for the  $D(s)$ , with  $s = 0.1\%, 0.2\%, \dots, 0.6\%$ . For scenario 2 where the parameter uncertainty is large, the average percentage deviation,  $\overline{PE}$ , is lower with the  $N \approx 64$  *SR* draws than with the  $N = 512$  *ESLP* draws and is 2.5, 1.3 and 1.1 times lower than with the  $N = 1024$  *PMC*, *MLHS* and the randomized Halton draws. In summary, for the 4-parameter case, the *SR* method performs extremely well compared to the other methods no matter whether the prior uncertainty is small or large. As the number of parameters increase (see Table 7 for the 8-parameter case), the *SR* approach is still much better than the others when the parameter uncertainty is small. However, the advantage of this approach decreases when the parameter uncertainty is large. It is seen that in Scenario 4, *SR* performs similarly to the *ESLP* approach for practically small sample sizes.

\*Figure 1\*

\*Figure 2\*

When the parameter uncertainty is small, for both 4- and 8-parameter cases, the Gaussian-Hermite quadrature closely follows (or is equally well as) the *SR* approach except in the case given in



Scenario 3, with  $N = 128$ . In that case, the Gaussian-Hermite quadrature approach performs badly. This is because this approach is poor if any of the dimensions is approximated with just one draw. The total number of draws  $N = 128$  is not large enough to ensure that all 8 parameters have at least two draws. However, once the number of draws is large enough, its performance improves dramatically (see the large change in performance in Scenario 3 from  $N = 128$  to  $N = 256$ ). When the parameter uncertainty is large, for the 4-parameter case, the *SR* approach is clearly better than the Gaussian-Hermite approach for small sample size  $N \approx 64$  and 128. However, the Gaussian-Hermite approach is still better than the other approaches. With a sample size which is larger than or equal to 256, the Gaussian method is the best. However, as the number of parameter increases, this approach is inefficient (see Table 7, Scenario 4).

For both 4- and 8-parameter cases, the *QMC* with the *ESLP* sequences, in general, outperforms that with the *MLHS* and the randomized Halton sequences. This holds in particular when the prior uncertainty is relatively small. When the parameter uncertainty is large, the advantage of the *ESLP* approach over the *MLHS* and the randomized Halton sequence becomes smaller. In addition, the performance of the *ESLP* sequences is comparable with and sometimes better than that of the *SR* in Scenario 4 where 8 parameters were considered and the prior uncertainty is large.

Finally, it seems that the randomized Halton sequences performs better than the *MLHS* approach when the total number of draws is equal or larger than 256. For smaller sample sizes, *MLHS* performs slightly better than the randomized Halton approach in most cases.

To investigate how the results depend on the designs that were chosen, we took 50 random designs for the two scenarios in the 4-parameter case. For each random design, we computed the  $\overline{PE}$  as before. Table 8 presents the averaged  $\overline{PE}$  values over the 50 random designs. These results support our previous conclusions.

\*Table 8\*

## 4 Conclusion and Discussion

In this study, we compared the performance of six sampling approaches in approximating the semi-Bayesian *D*-criterion in conjoint choice experiments. These sampling approaches include the simple *PMC* approach, three *QMC* approaches (*MLHS*, randomized Halton sequences and *ESLP* with Baker's transformation), the Gaussian-Hermite quadrature approach and the method based on a spherical-radial transformation. The comparison was conducted under four scenarios which include 4 and 8 parameters with small and large parameter uncertainty ( $\sigma = 0.2$  and  $\sigma = 1$ ). Based on the results, we can conclude that all methods lead to more accurate approximations under

small parameter uncertainty than under large parameter uncertainty. Therefore, when one uses a noninformative prior, more draws are needed to achieve the same accuracy as with an informative prior. In addition, it is shown that the *PMC* approach, which is commonly used for constructing experimental choice designs, is substantially less efficient than the other approaches in almost all cases. With this approach, a much larger sample size is required to achieve the same precision as with most of the other approaches. The *SR* method, based on the spherical-radial transformation, seems to be the most promising approach. In the 4 parameter cases, *SR* leads to significantly better results than the other methods no matter whether the parameter uncertainty is large or not. As the number of parameters increases to 8, *SR* is still much better than the others when the parameter uncertainty is small. However, as the parameter uncertainty becomes large, this approach has a similar performance to the *ESLP* approach for small samples sizes ( $N = 128, 256$ ). The performance of the Gaussian-Hermite quadrature depends on both the parameter uncertainty  $\sigma$  and the number of draws per parameter. It is very inefficient if the total number of draws  $N$  is too small to allow at least 2 draws for each dimension (see Scenario 3 with  $N = 128$ ). However, when  $\sigma$  is small, and as long as the number of draws for each parameter is at least two, the performance of the Gaussian-Hermite quadrature closely follows that from the *SR* approach. In this case, the *ESLP* with Baker's transformation is less efficient than the *SR* approach and the Gaussian-Hermite quadrature, but better than the *PMC*, *MLHS* and randomized Halton sequences. However, when  $\sigma$  is large, the Gaussian-Hermite quadrature is very inefficient in the 8 parameter case and the *ESLP* approach becomes more attractive.

So we can conclude that for small sample sizes (64, 128 and 256), the *SR* approach is in general, more efficient than the other methods. However, the drawback of this approach is that the number of draws grows fast with the number of parameters as the number of draws is  $N = N_R \times N_M \times (p+1)(p+2) + 1$ . As Gotwalt et al. (2008) suggested, at least two radii are required in the approximation for the *SR* approach working well. This implies that even with only one rotation ( $N_M = 1$ ), the number of draws required in the *SR* approach can be large for large  $p$ . For example, if  $p = 16$ , then *SR* requires at least 613 draws to be efficient in the approximation. In this situation, the *SR* approach is not an option used for small sample sizes. The Gaussian-Hermite quadrature approach has a similar problem. The total number of draws is computed by  $N = \prod_{k=1}^p N_k$ . It has been shown that the number of draws in each dimension  $N_k$  has to be at least 2 in order to provide reasonable efficiency. This means that with  $p = 16$ , at least 65536 draws should be used. Therefore, the Gaussian-Hermite quadrature and *SR* approaches are not practical when the dimension becomes large. As the number of draws of the *ESLP* with the Baker's transformation does not depend on  $p$ , this method seems most appropriate when the number of parameters is large.

## Acknowledgments

We are grateful to Zsolt Sándor for his helpful suggestions and contributions to this paper and to Xianni Lu for her help. She was funded by project NB/06/05 of the National Bank of Belgium. In addition, we acknowledge the financial support of the research council of the Katholieke Universiteit Leuven (OT-project 04/07).

## References

- Atkinson, A. C., Donev, A. N., 1992. Optimum experimental designs. Oxford U. K. Clarendon Press.
- Atkinson, A. C., Haines, L. M., 1996. Designs for non-linear and generalized linear models. In Handbook of Statistics, Volume 13, S. Ghosh and C. R. Rao (eds), 437- 475. New York: Elsevier Science.
- Ben-Akiva, M., Lerman, S., 1985. Discrete Choice Analysis. MIT Press, Cambridge, MA.
- Bhat, C. R., 2001. Quasi-random maximum simulated likelihood estimation of the mixed multinomial logit model. *Transportation Research Part B: Methodological*. 35 677-693.
- Bhat, C., 2003. Simulation estimation of mixed discrete choice models using randomized and scrambled Halton sequences. *Transportation Research Part B: Methodological*. 37 837-855.
- Bliemer, M. C. J., Rose, J. M., Hess, S., 2008. Approximation of Bayesian efficiency in experimental choice designs. *Journal of Choice Modeling*. Forthcoming.
- Brownstone, D., Train, K., 1999. Forecasting new product penetration with flexible substitution patterns. *Journal of Econometrics*. 89 109-129.
- Cassity C. R., Hopper, H. R., 1964. Abscissas and coefficients for the generalized Gauss-Laguerre quadrature formula using the zero ordinate. Nasa Technical Memorandum.
- Dror, H. A., Steinberg, D. M., 2006. Robust experimental design for multivariate generalized linear models. *Technometrics*. 48 520-529.
- Gotwalt, C. M., Jones, B. A., Steinberg, D. M., 2008. Fast computation of designs roust to parameter uncertainty for nonlinear settings. *Technometrics*. Forthcoming.
- Halton, J., 1960. On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals. *Numerische Mathematik*. 2 84-90.

- Hess, S., Train, K., Polak, J. W., 2006. On the use of a Modified Latin Hypercube Sampling (MLHS) method in the estimation of a Mixed Logit Model for vehicle choice. *Transportation Research B: Methodological*. 40 147-163.
- Hickernell F.J., Hong, H. S., L'Ecuyer, P., Lemieux, C., 2000. Extensible lattice sequences for quasi-Monte Carlo quadrature. *SIAM Journal on Scientific Computing*. 22 1117-1138.
- Huber, J., Zwerina, K., 1996. The importance of utility balance in efficient choice designs. *Journal of Marketing Research*. 33 307-317.
- Kessels, R., Goos, P., Vandebroek, M., 2006. A comparison of criteria to design efficient choice experiments. *Journal of Marketing Research*. 43 409-419.
- Kessels, R., Jones, B., Goos, P., Vandebroek, M., 2008. An efficient algorithm for constructing Bayesian optimal choice designs. *Journal of Business and Economic Statistics*. Forthcoming.
- Korobov, N.M., 1959. Approximate evaluation of repeated integrals. *Doklady Akademii Nauk SSSR*. 124 1207-1210.
- McFadden, D., 1974. Conditional logit analysis of qualitative choice behavior. In *Frontiers in Econometrics*, ed. Paul Zarembka: 105-142. New York: Academic Press.
- McFadden, D., 1978. Modeling the choice of residential location. In *Karquist, A. et al. (Eds.), Spatial Interaction Theory and Planning Models*. North-Holland, Amsterdam.
- Monahan J., Genz A., 1997. Spherical-radial integration rules for Bayesian computation. *Journal of the American Statistical Association*. 92 664-674.
- Mysovskikh, I. P., 1980. The approximation of multiple integrals using interpolatory cubature formulae in quantitative approximation. in *Quantitative Approximation*. eds. Devore, R. A., K. Sherer. New York: Academic Press. 217-243.
- Sándor, Z., András, P., 2004. Alternative sampling methods for estimating multivariate normal probabilities. *Journal of Econometrics*. 2 207-234.
- Sándor, Z., Train, K., 2004. Quasi-random simulation Of discrete choice models. *Transportation Research B*. 38 313-327.
- Sándor, Z., Wedel, M., 2001. Designing conjoint choice experiments using managers' prior beliefs. *Journal of Marketing Research*. 38 430-444.
- Sándor, Z., Wedel, M., 2002. Profile construction in experimental choice designs for mixed logit models. *Marketing Science*. 21 455-475.

- Sándor, Z., Wedel, M., 2005. Heterogeneous conjoint choice Designs. *Journal of Marketing Research*. 42 210-218.
- Stewart, G. W., 1980. The efficient generation of random orthogonal matrices with an application to condition estimation. *SIAM. Journal on Numerical Analysis*. 17 403-409.
- Train, K., 2000. Halton sequences for mixed logit. Working paper No. E00-278, Department of Economics, University of California, Berkeley.
- Tuffin, B., 1996. On the use of low-discrepancy sequences in Monte Carlo methods. *Monte Carlo Methods and Applications*. 2 295-320.
- Train, K., 2003. Discrete Choice Methods with Simulation. Cambridge University Press, New York.
- Vermeulen, B., Goos, P., Vandebroek, M., 2008. Models and optimal designs for conjoint choice experiments including a no-choice option. *International Journal of Research in Marketing*. 25 94-103.
- Yu, J., Goos, P., Vandebroek, M., 2008a. Model-robust design of conjoint choice experiments. *Communications in Statistics-Simulation and Computation*. 37 1603-1621.
- Yu, J., Goos, P., Vandebroek, M., 2008b. Efficient conjoint choice designs in the presence of respondent heterogeneity. *Marketing Science*. Forthcoming.
- Yu, J., Goos, P., Vandebroek, M., 2008c. A comparison of different Bayesian design criteria for constructing efficient conjoint choice experiments. Submitted for publication.

Table 1: Modified abscissas and weights for Gaussian-Hermite integration

$N$	$\tilde{a}_i^N$	$\tilde{w}_i^N$
1	0.0000000000	1.0000000000
2	$\pm 1.0000000000$	0.5000000000
3	0.0000000000 $\pm 1.7320508075$	0.6666666667 0.1666666667
4	$\pm 2.3344142184$ $\pm 0.7419637843$	0.0458758548 0.4541241452
5	0.0000000000 $\pm 2.8569700139$ $\pm 1.3556261799$	0.5333333333 0.0112574113 0.2220759220
6	$\pm 3.3242574335$ $\pm 1.8891758777$ $\pm 0.6167065902$	0.0025557844 0.0886157460 0.4088284696

Table 2: Radii and the weights for  $s = 1$

$N_R$	$i$	$r_i$	$w_{R_i}$
1	1	0.0000000000000000	0.3333333333333333
	2	1.732050807568877	0.6666666666666667
2	1	0.0000000000000000	0.1666666666666667
	2	2.0000000000000000	0.7500000000000000
	3	3.464101615137755	0.0833333333333333
3	1	0.0000000000000000	0.1000000000000000
	2	1.742060320814071	0.683737837696136
	3	2.936522819158578	0.210073649657705
	4	4.282763076707155	0.006188512646159
4	1	0.0000000000000000	0.0666666666666667
	2	1.566373686896139	0.591414023622233
	3	2.612472912279454	0.311625494160183
	4	3.715559878632455	0.029944275782052
	5	4.991600298892358	0.000349539768866

Table 4: Extended simplex points in 3 dimensions

1	1.000000	0.000000	0.000000
2	-0.333333	0.942809	0.000000
3	-0.333333	-0.471405	0.816497
4	-0.333333	-0.471405	-0.816497
5	-1.000000	0.000000	0.000000
6	0.333333	-0.942809	0.000000
7	0.333333	0.471405	-0.816497
8	0.333333	0.471405	0.816497
9	0.577350	0.816497	0.000000
10	0.577350	-0.408248	0.707107
11	0.577350	-0.408248	-0.707107
12	-0.577350	0.408248	0.707107
13	-0.577350	0.408248	-0.707107
14	-0.577350	-0.816497	0.000000
15	-0.577350	-0.816497	0.000000
16	-0.577350	0.408248	-0.707107
17	-0.577350	0.408248	0.707107
18	0.577350	-0.408248	-0.707107
19	0.577350	-0.408248	0.707107
20	0.577350	0.816497	0.000000

Table 3: Extended simplex points in 2 dimensions

1	1.000000	0.000000
2	-0.500000	0.866025
3	-0.500000	-0.866025
4	-1.000000	0.000000
5	0.500000	-0.866025
6	0.500000	0.866025
7	0.500000	0.866025
8	0.500000	-0.866025
9	-1.000000	0.000000
10	-0.500000	-0.866025
11	-0.500000	0.866025
12	1.000000	0.000000

Table 5: Summary of Scenarios

Scenarios	$p$	$\sigma$
1	4	0.2
2	4	1
3	8	0.2
4	8	1

Table 6: Evaluation of six sampling approaches in scenarios 1 and 2

N	Scenario 1: 4 parameters, $\sigma = 0.2$							Scenario 2: 4 parameters, $\sigma = 1$						
		<i>PMC</i>	<i>MLHS</i>	<i>Halt</i>	<i>ESLP</i>	<i>SR<sup>a</sup></i>	<i>Gau<sup>b</sup></i>		<i>PMC</i>	<i>MLHS</i>	<i>Halt</i>	<i>ESLP</i>	<i>SR<sup>a</sup></i>	<i>Gau<sup>b</sup></i>
64	$\overline{PE}$	1.11%	0.23%	0.29%	0.14%	0.01%	0.02%	$\overline{PE}$	7.16%	4.54%	4.40%	4.44%	0.82%	3.53%
	D(0.1%)	3	21	19	36	100	-	D(0.5%)	2	3	5	5	36	-
	D(0.2%)	10	49	48	79	100	-	D(1.0%)	10	10	11	12	65	-
	D(0.3%)	14	72	61	91	100	-	D(2.0%)	10	30	30	28	94	-
	D(0.4%)	23	83	80	99	100	-	D(3.0%)	26	40	45	44	100	-
	D(0.5%)	30	92	84	100	100	-	D(4.0%)	37	53	61	58	100	-
	D(0.6%)	33	96	93	100	100	-	D(5.0%)	46	66	70	67	100	-
128	$\overline{PE}$	0.71%	0.16%	0.20%	0.11%	0.01%	0.01%	$\overline{PE}$	5.36%	3.12%	3.06%	3.02%	0.70%	1.99%
	D(0.1%)	6	29	33	55	100	-	D(0.5%)	9	8	8	13	43	-
	D(0.2%)	16	63	61	85	100	-	D(1.0%)	14	18	15	24	78	-
	D(0.3%)	19	91	78	97	100	-	D(2.0%)	25	36	40	38	99	-
	D(0.4%)	31	97	87	100	100	-	D(3.0%)	34	50	59	60	100	-
	D(0.5%)	44	100	92	100	100	-	D(4.0%)	48	69	73	74	100	-
	D(0.6%)	49	100	97	100	100	-	D(5.0%)	57	87	81	85	100	-
256	$\overline{PE}$	0.52%	0.12%	0.10%	0.03%	0.01%	0.01%	$\overline{PE}$	3.62%	2.01%	1.95%	1.66%	0.62%	0.12%
	D(0.1%)	19	47	60	100	100	-	D(0.5%)	9	27	20	23	45	-
	D(0.2%)	27	80	85	100	100	-	D(1.0%)	16	41	30	39	78	-
	D(0.3%)	35	99	98	100	100	-	D(2.0%)	36	56	60	73	100	-
	D(0.4%)	40	100	100	100	100	-	D(3.0%)	49	76	80	84	100	-
	D(0.5%)	51	100	100	100	100	-	D(4.0%)	64	87	88	94	100	-
	D(0.6%)	55	100	100	100	100	-	D(5.0%)	76	92	95	96	100	-
512	$\overline{PE}$	0.37%	0.08%	0.06%	0.02%	0.00%	0.00%	$\overline{PE}$	2.46%	1.53%	1.28%	0.90%	0.52%	0.10%
	D(0.1%)	16	74	80	100	100	-	D(0.5%)	17	17	29	38	50	-
	D(0.2%)	34	99	100	100	100	-	D(1.0%)	32	35	47	65	92	-
	D(0.3%)	52	100	100	100	100	-	D(2.0%)	53	69	81	95	100	-
	D(0.4%)	60	100	100	100	100	-	D(3.0%)	67	91	95	99	100	-
	D(0.5%)	73	100	100	100	100	-	D(4.0%)	79	99	98	99	100	-
	D(0.6%)	82	100	100	100	100	-	D(5.0%)	86	99	99	99	100	-
1024	$\overline{PE}$	0.29%	0.06%	0.04%	0.01%	0.00%	0.00%	$\overline{PE}$	2.04%	1.07%	0.94%	0.65%	0.40%	0.10%
	D(0.1%)	21	84	96	100	100	-	D(0.5%)	13	22	40	50	54	-
	D(0.2%)	41	98	100	100	100	-	D(1.0%)	30	55	66	85	99	-
	D(0.3%)	60	100	100	100	100	-	D(2.0%)	57	89	88	95	100	-
	D(0.4%)	72	100	100	100	100	-	D(3.0%)	78	99	97	98	100	-
	D(0.5%)	90	100	100	100	100	-	D(4.0%)	89	100	99	99	100	-
	D(0.6%)	92	100	100	100	100	-	D(5.0%)	95	100	100	100	100	-

<sup>a</sup> For the *SR* approach, we used  $N = 61, 121, 241, 481, 1021$ .<sup>b</sup> For the Gaussian-Hermite quadrature approach, we used  $N = 64, 120, 256, 480, 1024$ .



Table 7: Evaluation of six sampling approaches in scenarios 3 and 4

N	Scenario 3: 8 parameters, $\sigma = 0.2$							Scenario 4: 8 parameters, $\sigma = 1$						
		<i>PMC</i>	<i>MLHS</i>	<i>Halt</i>	<i>ESLP</i>	<i>SR</i> <sup>a</sup>	<i>Gau</i> <sup>b</sup>		<i>PMC</i>	<i>MLHS</i>	<i>Halt</i>	<i>ESLP</i>	<i>SR</i> <sup>a</sup>	<i>Gau</i> <sup>b</sup>
128	$\overline{PE}$	0.57%	0.19%	0.24%	0.12%	0.01%	0.76%	$\overline{PE}$	4.46%	3.04%	3.09%	2.8%	2.09%	14.36%
	D(0.1%)	11	30	30	49	100	-	D(0.5%)	6	11	10	12	19	-
	D(0.2%)	23	54	54	81	100	-	D(1.0%)	12	26	24	25	31	-
	D(0.3%)	31	76	70	96	100	-	D(2.0%)	29	45	41	44	55	-
	D(0.4%)	41	92	78	100	100	-	D(3.0%)	45	55	55	65	71	-
	D(0.5%)	50	96	86	100	100	-	D(4.0%)	57	71	69	74	88	-
	D(0.6%)	57	100	92	100	100	-	D(5.0%)	67	80	81	87	96	-
256	$\overline{PE}$	0.36%	0.12%	0.12%	0.04%	0.01%	0.03%	$\overline{PE}$	3.03%	1.97%	1.93%	1.90%	1.81%	7.00%
	D(0.1%)	18	47	50	93	100	-	D(0.5%)	7	15	13	12	14	-
	D(0.2%)	37	86	85	100	100	-	D(1.0%)	20	28	32	30	27	-
	D(0.3%)	54	96	96	100	100	-	D(2.0%)	37	55	57	63	63	-
	D(0.4%)	65	98	99	100	100	-	D(3.0%)	58	66	79	85	81	-
	D(0.5%)	73	100	100	100	100	-	D(4.0%)	71	84	87	94	95	-
	D(0.6%)	78	100	100	100	100	-	D(5.0%)	84	93	95	95	99	-
512	$\overline{PE}$	0.26%	0.10%	0.08%	0.03%	0.00%	0.02%	$\overline{PE}$	2.35%	1.61%	1.34%	1.30%	1.57%	5.09%
	D(0.1%)	23	60	71	99	100	-	D(0.5%)	13	24	23	26	19	-
	D(0.2%)	49	90	95	100	100	-	D(1.0%)	25	41	47	45	37	-
	D(0.3%)	62	97	99	100	100	-	D(2.0%)	49	68	75	77	69	-
	D(0.4%)	79	99	100	100	100	-	D(3.0%)	73	82	95	95	91	-
	D(0.5%)	87	100	100	100	100	-	D(4.0%)	84	96	96	100	98	-
	D(0.6%)	93	100	100	100	100	-	D(5.0%)	88	98	100	100	100	-
1024	$\overline{PE}$	0.20%	0.07%	0.05%	0.02%	0.00%	0.01%	$\overline{PE}$	1.54%	1.04%	0.97%	0.92%	1.40%	3.89%
	D(0.1%)	33	77	89	100	100	-	D(0.5%)	25	38	31	42	25	-
	D(0.2%)	56	97	100	100	100	-	D(1.0%)	38	57	59	63	40	-
	D(0.3%)	74	100	100	100	100	-	D(2.0%)	70	83	90	91	70	-
	D(0.4%)	87	100	100	100	100	-	D(3.0%)	93	95	98	97	95	-
	D(0.5%)	94	100	100	100	100	-	D(4.0%)	97	99	99	100	99	-
	D(0.6%)	98	100	100	100	100	-	D(5.0%)	100	100	100	100	100	-

<sup>a</sup> For the *SR* approach, we used  $N = 181, 271, 541, 991$ .<sup>b</sup> For the Gaussian-Hermite quadrature approach, we used  $N = 128, 256, 576, 1152$ .

Table 8: Evaluation of six sampling approaches with 50 random designs in terms of  $\overline{PE}$  values for scenarios 1 and 2

Scenario 1						
N	<i>PMC</i>	<i>MLHS</i>	<i>Halt</i>	<i>ESLP</i>	<i>SR</i>	<i>Gau</i>
64	0.92%	0.21%	0.30%	0.12%	0.01%	0.01%
128	0.63%	0.14%	0.16%	0.09%	0.01%	0.01%
256	0.47%	0.09%	0.09%	0.03%	0.01%	0.01%
Scenario 2						
N	<i>PMC</i>	<i>MLHS</i>	<i>Halt</i>	<i>ESLP</i>	<i>SR</i>	<i>Gau</i>
64	6.94%	4.10%	4.12%	3.68%	0.59%	2.53%
128	4.68%	2.58%	2.53%	2.53%	0.44%	1.29%
256	3.40%	1.79%	1.73%	1.46%	0.34%	0.11%

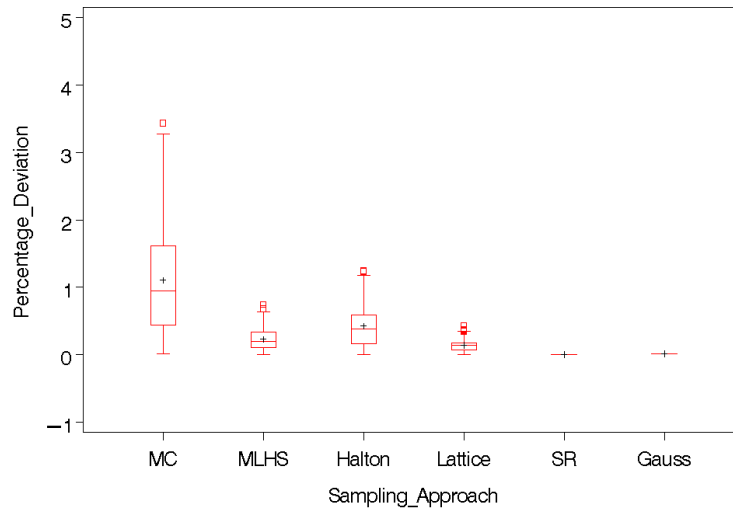


Figure 1: Percentage deviations of the approximated values from the true value in Scenario 1 with  $N = 64$

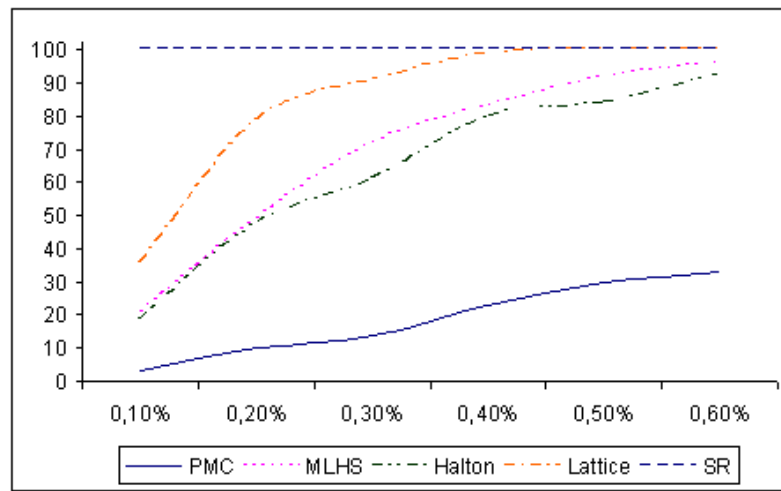


Figure 2: The number of times that the approximated values fall within a certain percentage of the true value in Scenario 1 with  $N = 64$